TWO DIMENSIONAL COMPRESSIVE CLASSIFIER FOR SPARSE IMAGES

Armin Eftekhari, Hamid Abrishami Moghaddam

K.N. Toosi University of Technology, Tehran, Iran

Massoud Babaie-Zadeh

Sharif University of Technology, Tehran, Iran

ABSTRACT

The theory of compressive sampling involves making random linear projections of a signal. Provided signal is sparse in some basis, small number of such measurements preserves the information in the signal, with high probability. Following the success in signal reconstruction, compressive framework has recently proved useful in classification, particularly hypothesis testing. In this paper, conventional random projection scheme is first extended to the image domain and the key notion of concentration of measure is closely studied. Findings are then employed to develop a 2D compressive classifier (2D-CC) for sparse images. Finally, theoretical results are validated within a realistic experimental framework.

1. INTRODUCTION

The recently developed theory of compressive sampling (CS) involves making random linear projections of a signal by multiplying a random matrix. Provided the signal is sparse in some basis, few random projections preserve the information in the signal, with high probability [1]. This remarkable result is rooted in the concentration of measure phenomenon [1], which implies that the Euclidean length of a vector is uniformly “shrunk” under a variety of random projection matrices, with high probability. Due to tangible advantages, CS framework has found many promising applications in signal and image acquisition, compression, and medical image processing [1, 2, 3]. The CS community, however, has mainly focused on the signal reconstruction problem from random projections to date [11] and other applications of CS have yet remained unexplored. In particular, few recent studies showed that classification can be accurately accomplished using random projections, which suggests random projections as an effective and reliable, yet universal feature extraction and dimension reduction tool. In this context, compressive detection, hypothesis testing, and manifold-aided classification of one-dimensional (1D) signals have been studied [1, 2]. However, as shown later in this paper, direct extension of these results to image domain (2D) is computationally prohibitive, which strongly hinders the application of conventional compressive classifier (1D-CC) in real-world scenarios. To overcome this major drawback, the idea of 2D compressive classification is developed in this paper. First, 2D random projection scheme is introduced in Section 2 and associated concentration properties are studied. It is then observed that Gaussian random matrices, as the most common choice in 1D compressive framework, are not appropriate for our 2D random projection scheme. Then, by adding an assumption of so-called 2D sparsity (in some basis) to images, desirable concentration properties are proved for the same set of admissible random

---

This research was supported in part by Iran Telecommunication Research Center.

matrices as in 1D framework. This assumption is not restrictive, as most images become sparse under well-known transformations, like DCT, or have a sparse edge map. In Section 3, these findings are exploited to develop a 2D compressive classifier (2D-CC) for sparse images, along with derivation of error bound for an important special case. Finally, 2D-CC is applied to retinal identification within a realistic setting. It is observed that, at worst, 2D-CC provides significant saving on computational load and memory requirements compared to 1D-CC, at the cost of negligible loss in performance. This performance loss, however, can be avoided by "wise" selection of parameters. We note that, due to space limitations, some intermediate steps of the derivations have been omitted, and the interested reader is referred to [4] for details.

2. CONCENTRATION OF MEASURE FOR 2D RANDOM PROJECTION SCHEME

In order to linearly project a given image $\mathrm{X} \in \mathbb{R}^{n_1 \times n_2}$ to lower dimensional space, columns of X are conventionally stacked into a vector $\mathrm{x} = \mathrm{vec}(\mathrm{X})$. This process, however, ignores the intrinsic row/column-wise structure of the image and, even for moderately sized matrices, involves prohibitive computational load and memory requirements for generation and manipulation of the projection matrix. As a remedy to these drawbacks, one may use so-called 2D projection scheme, i.e. $\mathrm{Y} = \mathrm{AXB}^{\mathrm{T}}$, in which $\mathrm{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathrm{A} \in \mathbb{R}^{m_1 \times n_1}$, $\mathrm{B} \in \mathbb{R}^{m_2 \times n_2}$, with $m_1 < n_1$ and $m_2 < n_2$. Note that, in terms of storage requirements, 2D projection scheme requires only $m_1 n_1 + m_2 n_2$ memory units for projection matrices, whereas conventional projection scheme to $\mathbb{R}^{m_1 m_2}$requires $m_1 m_2 n_1 n_2$ memory units. Similarly, using a naive matrix multiplication technique, computational complexity of 2D projection scheme is $\mathcal{O}(n^3)$, whereas conventional projection scheme to $\mathbb{R}^{m_1 m_2}$ has the complexity of order $\mathcal{O}(n^4)$. This section is devoted to the study of 2D projection scheme in the compressive framework. Clearly, successful inference in this scenario depends on preservation of the structure of samples after projection [1]. For 1D signals, this has received extensive treatment. Particularly, it has been shown that, given 1D signal, random matrices whose entries are i.i.d. random variables (rv's) with proper tail bounds, e.g. Gaussian or Bernoulli rv's, uniformly "shrink" the signal length after projection, with high probability. Using the union bound, one may then show that, with high probability, such random matrices preserve the structure of a set of samples after projection, by uniformly shrinking the pair-wise Euclidean distances. This, in turn, leads to optimal performance of statistical inference tasks in low-dimensional space. Focusing on Gaussian random matrices, as the most common choice in CS framework, a similar strategy is developed here.

Consider random Gaussian matrices $\mathrm{A} \in \mathbb{R}^{m_1 \times n_1}$ and $\mathrm{B} \in \mathbb{R}^{m_2 \times n_2}$, whose entries are i.i.d. $\mathcal{N}(0,1/n_1)$ and $\mathcal{N}(0,1/n_2)$ rv's, respectively. Given $\mathrm{X} \in \mathbb{R}^{n_1 \times n_2}$, let us define $\mathrm{x} = \mathrm{vec}(\mathrm{X})$ and $\mathrm{D} = \mathrm{B} \otimes \mathrm{A}$, where $\otimes$ denotes the Kronecker product. Thus, $\mathrm{AXB}^{\mathrm{T}} = \mathrm{Dx}$, where each entry of D has the following distribution.

$$p(D_{i,j}) = \frac{n_1 n_2}{\pi} K_0\left(n_1 n_2 |D_{i,j}|\right) \tag{1}$$

in which $K_0(\cdot)$ denotes the modified Bessel function of second kind. We are interested in studying the concentration of rv $\|\mathrm{Dx}\|^2$ about its expected value $(m_1 m_2 / n_1 n_2)\|\mathrm{x}\|^2$ by finding an exponentially-fast decreasing bound of the form:

$$\Pr\left[\left|\|\mathrm{Dx}\|^2 - \frac{m_1 m_2}{n_1 n_2}\|\mathrm{x}\|^2\right| \geq \frac{m_1 m_2}{n_1 n_2}\epsilon\right] \leq e^{-c(\epsilon) m_1 m_2} \tag{2}$$

where probability is taken over all matrices A and B, and $c(\epsilon)$ depends only on ϵ. Note that (in contrast to 1D counterpart), entries of D are no more independently distributed. Furthermore, each entry of D is a product of two i.i.d. Gaussian rv's. Properties of the Kronecker product implies that each row of D is dependent with exactly $m_1 + m_2 - 2$ other rows, and that we can partition the rows of D into m nonoverlapping partitions $\{R_i\}_{i=1}^{m}$ with $|R_i| = m_2$, such that rows in each partition are independent. Let us denote by D_{R_i} the $m_1 \times n_1 n_2$ submatrix obtained by retaining the rows of D corresponding to the indices in R_i. Obviously, rows of D_{R_i} are independent. In addition, we have:

$$\|Dx\|^2 = \sum_{i=1}^{q} \|D_{R_i}x\|^2 \tag{4}$$

where entries of D_{R_i} are zero-mean rv's with variance $1/n_1 n_2$ and are distributed according to (1). Note that, for a given D, concentration of D_{R_i} implies that of D, i.e. the following statement holds:

$$\left\{\bigcap_{i=1}^{q}\left|\|D_{R_i}x\|^2 - \frac{|R_i|}{n_1 n_2}\|x\|^2\right| \geq \frac{|R_i|}{n_1 n_2}\epsilon\right\} \longrightarrow \left\{\left|\sum_{i=1}^{q}\|D_{R_i}x\|^2 - \sum_{i=1}^{q}\frac{|R_i|}{n_1 n_2}\|x\|^2\right| \geq \sum_{i=1}^{q}\frac{|R_i|}{n_1 n_2}\epsilon\right\} \tag{5}$$

which, after simplification of the right hand side, reduces to:

$$\left\{\bigcap_{i=1}^{q}\left|\|D_{R_i}x\|^2 - \frac{|R_i|}{n_1 n_2}\|x\|^2\right| \geq \frac{|R_i|}{n_1 n_2}\epsilon\right\} \longrightarrow \left\{\left|\|Dx\|^2 - \frac{m_1 m_2}{n_1 n_2}\|x\|^2\right| \geq \frac{m_1 m_2}{n_1 n_2}\epsilon\right\} \tag{6}$$

Therefore, we have:

$$\Pr\left[\bigcap_{i=1}^{q}\left|\|D_{R_i}x\|^2 - \frac{|R_i|}{n_1 n_2}\|x\|^2\right| \geq \frac{|R_i|}{n_1 n_2}\epsilon\right] \leq \Pr\left[\left|\|Dx\|^2 - \frac{m_1 m_2}{n_1 n_2}\|x\|^2\right| \geq \frac{m_1 m_2}{n_1 n_2}\epsilon\right] \tag{7}$$

or equivalently:

$$\Pr\left[\left|\|Dx\|^2 - \frac{m_1 m_2}{n_1 n_2}\|x\|^2\right| \leq \frac{m_1 m_2}{n_1 n_2}\epsilon\right] \leq \sum_{i=1}^{q}\Pr\left[\left|\|D_{R_i}x\|^2 - \frac{|R_i|}{n_1 n_2}\|x\|^2\right| \leq \frac{|R_i|}{n_1 n_2}\epsilon\right] \tag{8}$$

Now, under the assumption of strong concentration of $D_{R_i} \in \mathbb{R}^{|R_j| \times n_1 n_2}$ about its expected value, we may write:

$$\Pr\left[\left|\|Dx\|^2 - \frac{m_1 m_2}{n_1 n_2}\|x\|^2\right| \leq \frac{m_1 m_2}{n_1 n_2}\epsilon\right] \leq \sum_{i=1}^{q} e^{-c(\epsilon)|R_i|} \leq e^{-c(\epsilon)\left\lfloor\frac{m_1 m_2}{q}\right\rfloor + \ln q} \tag{9}$$

In other words, strong concentration of $\|D_{R_i}x\|^2$ implies that of $\|Dx\|^2$. For ease of notation, let $M = D_{R_i}$ and $R = |R_i|$. Now, it suffices to find an exponentially-fast decreasing bound for:

$$\Pr\left[|\|Mx\|^2 - \frac{R}{n_1}| \geq \frac{R}{n_1 n_2}\epsilon\right] \tag{10}$$

where, due to linearity in $\|x\|$, we have assumed that each column of X has unit length. Consider the probability of $\|Mx\|^2 - R/n_1 > R\epsilon/n_1 n_2$. Invoking the Chernoff bounding technique, for any $h > 0$, this probability can be written as:

$$\Pr\left[\|Mx\|^2 - \frac{R}{n_1} \geq \frac{R}{n_1 n_2}\epsilon\right] = \Pr\left[e^{h\|Mx\|^2} \geq e^{\frac{R}{n_1 n_2}h(\epsilon + n_2)}\right] \leq \mathbb{E}\left[e^{h\|Mx\|^2}\right]e^{-\frac{R}{n_1 n_2}h(\epsilon + n_2)} \tag{11}$$

where we have used the Markov's inequality. Since rows of M are independent, we have:

$$\mathbb{E}\left[e^{h\|Mx\|^2}\right] = \mathbb{E}\left[e^{h\sum_{i=1}^{R}\left(\sum_{j=1}^{n_1 n_2} M_{i,j}x_j\right)^2}\right] = \prod_{i=1}^{R}\mathbb{E}\left[e^{h\left(\sum_{j=1}^{n_1 n_2} M_{i,j}x_j\right)^2}\right] = \mathbb{E}\left[e^{h\left(\sum_{j=1}^{n_1 n_2} M_{1,j}x_j\right)^2}\right]^R \tag{12}$$

where, without any loss in generality, we have used the first row of M in (12). In fact, we may consider only the first row of D, i.e. $\mathrm{D}_{r_1} = [B_{1,1}\mathrm{A}_{r_1}\ B_{1,2}\mathrm{A}_{r_1} \dots B_{1,n_2}\mathrm{A}_{r_1}]$ in our analysis, where A_{r_1} denotes the first row of A. Denoting the jth column of X by X_{c_j}, we have:

$$\sum_{j=1}^{n_1 n_2} M_{1,j} x_j = B_{1,1}\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_1} + \cdots + B_{1,n_2}\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_{n_2}} \tag{13}$$

Therefore, the last expectation in (12) can be expanded as:

$$\mathbb{E}\left[\mathrm{e}^{h\left(\sum_{j=1}^{n_1 n_2} M_{1,j}x_j\right)^2}\right] = \sum_{k=0}^{\infty} \frac{h^k}{k!}\mathbb{E}\left[\left(\sum_{j=1}^{n_1 n_2} M_{1,j}x_j\right)^{2k}\right] \tag{14}$$

in which,

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{j=1}^{n_1 n_2} M_{1,j}x_j\right)^{2k}\right] &= \mathbb{E}\left[\left(B_{1,1}\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_1} + \cdots + B_{1,n_2}\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_{n_2}}\right)^{2k}\right] \\
&= \sum_{\substack{i_1+\cdots+i_{n_2}=2k \\ i_1,\dots,i_{n_2}\geq 0}} \frac{2k!}{i_1!\dots i_{n_2}!}\mathbb{E}\left[\left(B_{1,1}\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_1}\right)^{i_1}\dots\left(B_{1,n_2}\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_{n_2}}\right)^{i_{n_2}}\right] \\
&= \sum_{\substack{i_1+\cdots+i_{n_2}=2k \\ i_1,\dots,i_{n_2}\geq 0}} \frac{2k!}{i_1!\dots i_{n_2}!}\mathbb{E}\left[B_{1,1}^{i_1}\right]\dots\mathbb{E}\left[B_{1,n_2}^{i_{n_2}}\right]\mathbb{E}\left[\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_1}\right)^{i_1}\dots\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_{n_2}}\right)^{i_{n_2}}\right]
\end{aligned} \tag{15}$$

where in the last line we have used the fact that entries of A and B are i.i.d. rv's. Furthermore, entries of B are i.i.d. $\mathcal{N}(0,1/n_2)$ rv's. Hence, setting even moments to zero and replacing the odd moments with their value, (15) reduces to:

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{j=1}^{n_1 n_2} M_{1,j}x_j\right)^{2k}\right] &= \sum_{\substack{j_1+\cdots+i_{n_2}=k \\ j_1,\dots,j_{n_2}\geq 0}} \frac{2k!}{2j_1!\dots 2j_{n_2}!}\mathbb{E}\left[B_{1,1}^{2j_1}\right]\dots\mathbb{E}\left[B_{1,n_2}^{2j_{n_2}}\right]\mathbb{E}\left[\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_1}\right)^{2j_1}\dots\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_{n_2}}\right)^{2j_{n_2}}\right] \\
&= \sum_{\substack{j_1+\cdots+i_{n_2}=k \\ j_1,\dots,j_{n_2}\geq 0}} \frac{2k!}{2j_1!\dots 2j_{n_2}!}\left(\frac{1}{n_2}\right)^k \frac{2j_1!}{2^{j_1}j_1!}\cdots\frac{2j_{n_2}!}{2^{j_{n_2}}j_{n_2}!}\mathbb{E}\left[\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_1}\right)^{2j_1}\dots\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_{n_2}}\right)^{2j_{n_2}}\right]
\end{aligned} \tag{16}$$

Employing the Cauchy-Schwarz inequality and noting that columns of X has unit length, we may write:

$$\mathbb{E}\left[\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_1}\right)^{2j_1}\dots\left(\mathrm{A}_{r_1}^{\mathrm{T}}\mathrm{X}_{c_{n_2}}\right)^{2j_{n_2}}\right] \leq \mathbb{E}\left[\left\|\mathrm{A}_{r_1}\right\|^{2k}\right] = \mathbb{E}\left[\left(A_{1,1}^2 + \cdots A_{1,n_1}^2\right)^k\right] \tag{17}$$

Since entries of A are i.i.d. $N(0, 1/n_1)$, the scaled version of the sum on the right has a Chi-square with degree n_1. Therefore, using the moments of Chi-square distribution, we have:

$$\mathbb{E}\left[\left(A_{1,1}^2 + \cdots A_{1,n_1}^2\right)^k\right] = \left(\frac{1}{n_1}\right)^k 2^k \frac{\Gamma\left(k + \frac{n_1}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)} \tag{18}$$

where $\Gamma(\cdot)$ denotes the Gamma function. Finally, we may now simplify (16) to obtain:

$$\mathbb{E}\left[\left(\sum_{j=1}^{n_1 n_2} M_{1,j}x_j\right)^{2k}\right] \leq \sum_{\substack{j_1+\cdots+i_{n_2}=k \\ j_1,\ldots,j_{n_2}\geq 0}} \frac{2k!}{2j_1!\ldots 2j_{n_2}!}\frac{2j_1!}{2^{j_1}j_1!}\cdots\frac{2j_1!}{2^{j_1}j_1!}\left(\frac{2}{n_1 n_2}\right)^k \frac{\Gamma\left(k+\frac{n_1}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)}$$

$$= \left(\frac{1}{n_1 n_2}\right)^k \frac{\Gamma\left(k+\frac{n_1}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)}\frac{2k!}{k!} \sum_{\substack{j_1+\cdots+i_{n_2}=k \\ j_1,\ldots,j_{n_2}\geq 0}} \frac{k!}{j_1!\ldots j_{n_2}!} \tag{19}$$

$$= \left(\frac{2}{n_1 n_2}\right)^k \frac{2k!}{k!}\frac{\Gamma\left(k+\frac{n_1}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)}$$

This result let us rewrite (14) as follows:

$$\mathbb{E}\left[e^{h\left(\sum_{j=1}^{n_1 n_2} M_{1,j}x_j\right)^2}\right] \leq \sum_{k=1}^{\infty} \frac{2k!}{(k!)^2}\left(\frac{2h}{n_1 n_2}\right)^k \frac{\Gamma\left(k+\frac{n_1}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)} \tag{20}$$

which clearly fails to converge. Consequently, the probability of $\|\mathrm{Mx}\|^2 - R/n_1 \geq R\epsilon/n_1 n_2$ fails to decrease exponentially fast, in general; not letting us to extend the random projection scheme to image domain for random Gaussian matrices. In the following, rather than using a distribution with weaker concentration around the origin, we prefer to insert an additional constraint on X, which would allow for successful operation of 2D random projection scheme with the same set of admissible random matrices as in 1D case. We shall first start with a brief review on 1D case. Let Σ_k denote the set of all signals of length n with at most k nonzero entries. We say that $\mathrm{A} \in \mathbb{R}^{m\times n}$ satisfies restricted isometry property (RIP) of order k, if for every $\mathrm{v} \in \Sigma_k$, the following holds for some $\epsilon_k \in [0,1)$.

$$(1-\epsilon_k)\frac{m}{n}\|\mathrm{v}\|^2 \leq \|\mathrm{Av}\|^2 \leq \|\mathrm{v}\|^2\frac{m}{n}(1+\epsilon_k) \tag{21}$$

Using arguments on tail bounds, it is shown that many random matrices satisfy RIP condition with high probability [1, 3]. For instance, $m \times n$ matrices whose entries are i.i.d. $\mathcal{N}(0,1/n)$ rv's satisfy RIP of order k with probability exceeding $1-e^{-c_2(\epsilon_k)m}$, provided $k \leq c_1(\epsilon_k)m \log n/k$, where $c_1(\epsilon_k)$ and $c_2(\epsilon_k)$ depend only on ϵ_k [1, 3]. Also random orthoprojectors, i.e. matrices with random orthonormal rows, satisfy the RIP condition similarly. Extending these ideas to image domain, let Σ_{k_r,k_c} denote the set of all images $\mathrm{V} \in \mathbb{R}^{n_1\times n_2}$, whose nonzero entries are distributed in at most k_r rows and k_c columns. A matrix with this property will be called 2D sparse [3]. Now, we can make the following observation.

Observation 1. Suppose projection matrices $\mathrm{A} \in \mathbb{R}^{m_1\times n_1}$ and $\mathrm{B} \in \mathbb{R}^{m_2\times n_2}$, respectively, satisfy the RIP conditions of orders k_r and k_c for some $\epsilon_{2k_r}, \epsilon_{2k_c} \in [0,1)$. Then, for any $\mathrm{X} \in \Sigma_{2k_r,2k_c}$, we have:

$$(1-\epsilon_{2k_r})(1-\epsilon_{2k_c})\frac{m_1 m_2}{n_1 n_2}\|\mathrm{X}\|^2 \leq \|\mathrm{AXB^T}\|^2 \leq \|\mathrm{X}\|^2\frac{m_1 m_2}{n_1 n_2}(1+\epsilon_{2k_r})(1+\epsilon_{2k_c}) \tag{22}$$

Proof. Without loss of generality, assume that $\mathrm{X} = \begin{bmatrix}\underline{\mathrm{X}} & 0\\ 0 & 0\end{bmatrix}$, where $\underline{\mathrm{X}}$ has at most $2\mathrm{k_r}$ rows and $2\mathrm{k_c}$ columns. Then, using the well-known properties of the Kronecker product, we may write:

$$|\mathrm{AXB^T}|^2 = \underline{\mathrm{x}}^\mathrm{T}(\underline{\mathrm{B}}^\mathrm{T}\underline{\mathrm{B}})\otimes(\underline{\mathrm{A}}^\mathrm{T}\underline{\mathrm{A}})\underline{\mathrm{x}} \tag{23}$$

where $\underline{\mathrm{x}} = \mathrm{vec}(\underline{\mathrm{X}})$, and $\underline{\mathrm{A}}$ and $\underline{\mathrm{B}}$ are column submatrices of A and B with at most $2k_r$ and $2k_c$ columns, respectively. Then, clearly, range of $|\mathrm{AXB^T}|^2/|\mathrm{X}|^2$ is equal to the range of eigenvalues of $(\underline{\mathrm{B}}^{\mathrm{T}}\underline{\mathrm{B}})\otimes(\underline{\mathrm{A}}^{\mathrm{T}}\underline{\mathrm{A}})$. Now, note that the range of eigenvalues of $(\underline{\mathrm{B}}^{\mathrm{T}}\underline{\mathrm{B}})\otimes(\underline{\mathrm{A}}^{\mathrm{T}}\underline{\mathrm{A}})$ is the product of the ranges of eigenvalues of $\underline{\mathrm{B}}^{\mathrm{T}}\underline{\mathrm{B}}$ and $\underline{\mathrm{A}}^{\mathrm{T}}\underline{\mathrm{A}}$ [6]. On the other hand, RIP conditions on A and B imply that the range of eigenvalues of $(\underline{\mathrm{B}}^{\mathrm{T}}\underline{\mathrm{B}})\otimes(\underline{\mathrm{A}}^{\mathrm{T}}\underline{\mathrm{A}})$ is $\frac{m_1 m_2}{n_1 n_2}\left(1 \pm \epsilon_{2k_c}\right)\left(1 \pm \epsilon_{2k_r}\right)$. Our claim follows. ■

Therefore, if A and B are admissible random matrices in the 1D compressive framework, like Gaussian or Bernouuli random matrices, then (5) holds with probability exceeding $1 - e^{-c_2(\epsilon_{2k_r})m_1 - c_2(\epsilon_{2k_c})m_2}$, provided $2k_r \le c_1(\epsilon_{2k_r})m_1 \log n_1/2k_r$ and $2k_c \le c_1(\epsilon_{2k_c})m_2 \log n_2/2k_c$. This means that, if $\mathrm{X} \in \Sigma_{2k_r,2k_c}$, then $\|\mathrm{AXB^T}\|^2/\|\mathrm{X}\|^2$ is strongly concentrated about its expected value for a wide range of random matrices; hence establishing the concentration of measure in 2D random projection scheme.

3. 2D COMPRESSIVE CLASSIFIER FOR SPARSE IMAGES

In many applications, images are either sparse in the pixel domain or have a sparse representation in some basis, such that their nonzero entries are concentrated in a small number of rows/columns. Examples include a images with sparse edge map or DCT transform of natural images [10]. For this class of images, obtained results in previous section are applicable; enabling us to develop 2D-CC. Formally, assume that $\mathcal{X} = \{\mathrm{X}_1, \dots, \mathrm{X}_L\}$ denotes a set of $n_1 \times n_2$ known 2D sparse images, such that $\mathrm{X}_l \in \Sigma_{k_r,k_c}$, $l = 1, \dots, L$, for some integers $k_r < n_1$, $k_c < n_2$. The (possibly noisy) "true" image $\mathrm{X}_T \in \mathcal{X}$ undergoes 2D random projection to obtain $\mathrm{Y} = \mathrm{A}(\mathrm{X}_T + \mathrm{N})\mathrm{B}^{\mathrm{T}}$, where $\mathrm{N} \in \mathbb{R}^{n_1 \times n_2}$ represents the noise. Now, we will be concerned with discrimination among the members of $\mathcal{X}$, given only low-dimensional random projections. Given A and B, failure will be quantified in terms of expected error. Let $\mathrm{y} = \mathrm{vec}(\mathrm{Y})$, $\mathrm{x} = \mathrm{vec}(\mathrm{X})$, $\mathrm{n} = \mathrm{vec}(\mathrm{N})$ and $\mathrm{D} = \mathrm{B}\otimes\mathrm{A}$. For simplicity of analysis, we further assume $\mathrm{n} \sim \mathcal{N}\left(0, \sigma^2 \mathrm{I}_{n_1 n_2}\right)$, where I_a denotes the $a \times a$ identity matrix. In addition, A and B are selected to be orthoprojectors, which implies that the distribution of noise remains unchanged under projection. Now, provided X_l's happen equally likely, the Bayes decision rule and the associated expected error would be:

$$\hat{\mathrm{x}}_l = \underset{\mathrm{x}_l \in \mathrm{vec}(\mathcal{X})}{\mathrm{argmin}} \|\mathrm{y} - \mathrm{Dx}_l\| = \underset{\mathrm{x}_l \in \mathcal{X}}{\mathrm{argmin}} \|\mathrm{Y} - \mathrm{AX}_l\mathrm{B}^{\mathrm{T}}\| \tag{24}$$

$$\mathrm{Err}(\mathrm{A}, \mathrm{B}) = 1 - \frac{1}{L}\int_{\mathrm{y}} \max_{l}\{\mathrm{p}_l(\mathrm{y})\}\, \mathrm{dy} \tag{25}$$

where $p_l(\mathrm{y}) = \mathcal{N}\left(\mathrm{Dx_l}, \sigma^2 \mathrm{I}_{n_1 n_2}\right)$ [9]. Note that, for any nonnegative sets $\{a_l\}$ and $\{s_l\}$ with $\sum_1^L s_l = 1$, we have $\max_l\{a_l\} \ge \prod_{l=1}^L a_l^{s_l}$. Consequently, we may write:

$$\mathrm{Err}(\mathrm{A}, \mathrm{B}) \le 1 - \frac{1}{L}\int_{\mathrm{y}} \prod_{l=1}^{L} \left(p_l(\mathrm{y})\right)^{s_l} \mathrm{dy} \tag{26}$$

Noting that $p_l(\mathrm{y}) = \mathcal{N}\left(\mathrm{Dx}_l, \sigma^2 \mathrm{I}_{m_1 m_2}\right)$, we have:

$$\prod_{l=1}^{L}\left(p_l(\mathrm{y})\right)^{s_l} = (2\pi\sigma^2)^{-\frac{m_1m_2}{2}} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{l=1}^{L} s_l\|\mathrm{y}-\mathrm{Dx}_l\|^2\right) \tag{27}$$

where the exponent can be re-written as:

$$\sum_{l=1}^{L} s_l\|\mathrm{y}-\mathrm{Dx}_l\|^2 = \left\|\mathrm{y}-\mathrm{D}\sum_{l=1}^{L} s_l\mathrm{x}_l\right\|^2 - \left\|\mathrm{D}\sum_{l=1}^{L} s_l\mathrm{x}_l\right\|^2 + \sum_{l=1}^{L} s_l\|\mathrm{Dx}_l\|^2 \tag{28}$$

Therefore, (26) may be written as:

$$\begin{aligned}\mathrm{Err}(\mathrm{A},\mathrm{B}) &\le 1 - \overbrace{\int (2\pi\sigma^2)^{-\frac{m_1m_2}{2}} \exp\left(-\frac{\|\mathrm{y}-\mathrm{D}\sum_{l=1}^{L} s_l\mathrm{x}_l\|^2}{2\sigma^2}\right)dy}^{1} \times \exp\left(\frac{\|\sum_{l=1}^{L}\ s_l\mathrm{Dx}_l\|^2 - \sum s_l\|\mathrm{Dx}_l\|^2}{2\sigma^2}\right) \\ &= 1 - \exp\left(\frac{\|\sum_{l=1}^{L}\ s_l\mathrm{Dx}_l\|^2 - \sum_{l=1}^{L} s_l\|\mathrm{Dx}_l\|^2}{2\sigma^2}\right)\end{aligned} \tag{29}$$

Though minimizing the bound in (29) with respect to s_l is straightforward, $s_l = 1/L$ is assumed for convenience, $l = 1, \dots, L$. With this choice, (29) simplifies to:

$$\mathrm{Err}(\mathrm{A},\mathrm{B}) \le 1 - \exp\left(\frac{\frac{\|\sum \mathrm{Dx}_l\|^2}{L^2} - \frac{\sum\|\mathrm{Dx}_l\|^2}{L}}{2\sigma^2}\right) = 1 - \exp\left(-\frac{\sum_{l\neq l'}\|\mathrm{D}(\mathrm{x}_l - \mathrm{x}_{l'})\|^2}{4\sigma^2L^2}\right) \tag{30}$$

Now, assuming that A and B satisfy the RIP conditions similar to those mentioned in Observation 1, and defining $d_{min} \triangleq \min_{l\neq l'}\|x_l - x_{l'}\|^2$, Observation 1 implies the following bound for classification error:

$$1 - \exp\left(-(1+\epsilon_{2k_r})(1+\epsilon_{2k_c})\frac{m_1m_2}{n_1n_2}\frac{(L-1)}{8\sigma^2L}d_{min}\right) \tag{31}$$

In particular, if A and B are random orthoprojectors, then above bound holds with conditions noted right after Observation 1. It is observed that the classification error decays exponentially fast as number of observations $\mathrm{m}_1\mathrm{m}_2$ increases.

4. EXPERIMENTS

In this section, the efficacy of the proposed 2D-CC is examined in retinal identification problem. Retinal biometrics refers to identity verification of individuals based on their retinal images. Salient anatomical features of retina are depicted in Fig. 1.a, among which vessel tree pattern is a superior biometric trait, as it is unique, time invariant and almost impossible to forge. Our experiments are conducted on VARIA database containing 153 (multiple) retinal images of 59 individuals [7]. To compensate for the variations in the location of optic disc (OD) in retinal images and to exploit the larger diameter of vessels near OD, a circular region of interest (ROI) in the vicinity of OD is used to construct the feature matrix. To extract the ROI, center of OD is first localized using template matching as in [8], followed by simple analysis of area. Then, similar to [8], vessel tree is extracted by a combination of local contrast enhancement and histogram thresholding. Then, a ring-shaped mask with proper radii centered at OD is used to form the feature matrix $\mathrm{X} \in \mathbb{R}^{n_1\times n_2}$ by collecting the pixels along $n_2 = 200$ beams of length $n_1 = 100$ originating from OD (Fig. 1.c). Note that feature matrices readily satisfy the requirements of Observation 1 in pixel domain, as $k_r < n_1$ and $k_c \ll n_2$. Due to small number of images per subject (~2) and approximate invariance of feature matrix to the location of OD, feature matrices of the same subject are modeled as noisy deviations from corresponding mean feature matrix. Hence, once all images are processed, $\mathcal{X} = \{X_1, \ldots, X_L\}$ is formed, where each X_l is the mean feature matrix of lth subject. Dimension reduction and classification of a test feature matrix is then performed in $\mathbb{R}^{m_1 \times m_2}$ by the proposed 2D-CC, and in $\mathbb{R}^{m_1 m_2}$ by 1D-CC. Error is measured using the leave-one-out scheme and the average results over 100 independent repetitions are depicted in Fig. 2 for a wide range of m_1 and m_2. Although explicit calculation of the bound in (31) is intractable [1], we note that the exponential nature of error is in accordance with our findings. Also, due to highly redundant nature of feature matrices along their columns, "wise" choices for m_1 and m_2, which consider this redundancy, exhibit good performance especially for small values of m_1 and m_2. In contrast, "careless" choices for m_1 and m_2, degrade the performance (Fig. 3). Using an Intel Core 2 Duo, 2.67 GHz processor with 3.24 GB of memory, we found that each repetition of 2D-CC approximately took $0.0098\, m_1 m_2$ seconds, whereas this number was roughly $0.655 m_1 m_2$ for 1D-CC, in MATLAB7 environment. Note that this difference is significant even with our small-sized feature matrices. In sum, for typical choices of m_1 and m_2, 2D-CC runs much faster than 1D-CC, yet producing results with negligible loss in performance. This loss, however, disappears with proper choice for m_1 and m_2 which takes the prior knowledge into account. In addition, 2D-CC enjoys significantly less memory requirements.

5. CONCLUSIONS

In this paper, the idea of random projections is extended to image domain and associated concentration properties are studied. Our findings are then used to develop 2D-CC, along with error bound for an important special case. Finally, results are validated in a realistic application.

6. REFERENCES

[1] M.A. Davenport, M.F. Duarte, M.B. Wakin, J.N. Laska, "The smashed filter for compressive classification and target recognition," Proc. SPIE Computational Imaging V, 2007.

[2] J. Haupt, R. Castro, R. Nowak, G. Fudge, A. Yeh, "Compressive sampling for signal classification," Fortieth Asilomar Conf. on Signals, Systems and Computers, 2006.

[3] A. Ghaffari, M. Babaie-Zadeh, and C. Jutten, "Sparse decomposition of two dimensional signals," IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'09), Accepted.

[4] J. Haupt, W.U. Bajwa, G. Raz, "Toeplitz compressed sensing matrices with applications to channel sensing," Submitted, 2008.

[5] A. Hajnal, and E. Szemeredi, "Proof of a Conjecture of P. Erdos," in Combinatorial Theory and its Application, P. Erdos, A. R´enyi, and V. T. S´os, Eds., pp. 601–623. North-Holland, Amsterdam, 1970.

[6] K.B. Peterson, M.S. Pederson, "Matrix Cookbook," available online at http://matrixcookbook.com

[7] VARIA database, available online at http://www.varpa.es/varia.html.

[8] H. Farzin, H. Abrishami, "A novel retinal identification system," EURASIP Journal on Advances in Signal Processing, 2008.

[9] R.O. Duda, P.E. Hart, D.G. Stork, "Pattern Classification," Jhon Wiley, 2nd edition, 2001.
[10] J.-L. Starck, M. Elad, D.L. Donoho, "Image Decomposition Via the Combination of Sparse Representations and a Variational Approach," IEEE Transactions on Image Processing, 2006.
[11] A. Eftekhari, M. Babaie-Zadeh, C. Jutten, H. Abrishami Moghaddam, "Robust-SL0 for stable sparse representation in noisy settings," IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'09), Accepted.

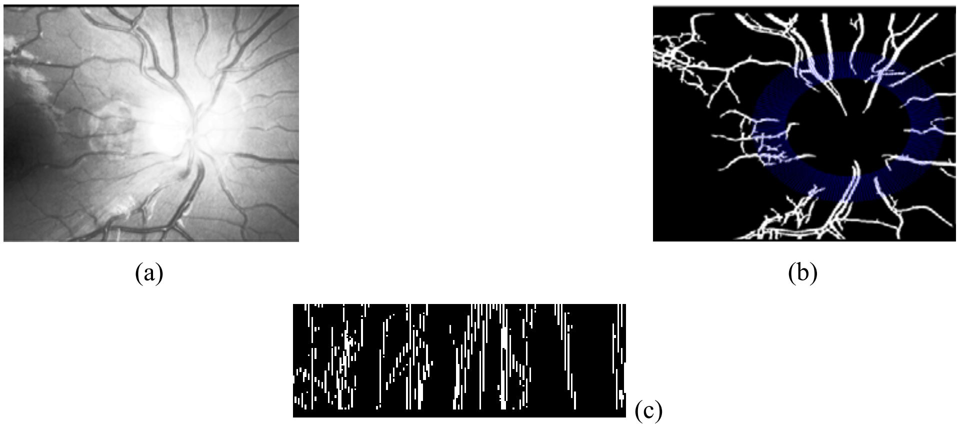

Fig. 1. (a) Retinal image; bright area is OD. (b) Vessel tree (in white) and mask (in blue). (c) Feature matrix for $n_1 = 100$, $n_2 = 300$. Due to limited space, images (a) and (b) are cropped.

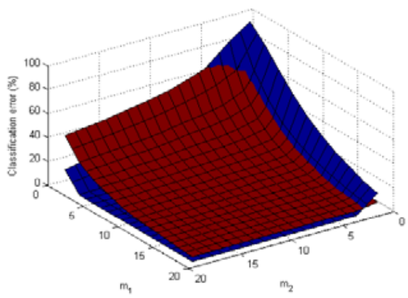

Fig. 2. Average classification error of 2D-CC (red surface) and 1D-CC (blue surface) for a wide range of m_1 and m_2.

Fig. 3. Two examples of "wise" choices which consider the redundancy along columns: $m_1 = 1$ (left) and $m_1 = 3$ (right)